The PASCAL Visual Object Classes Challenge 2009 (VOC2009)

Part 2 – Classification Task

Mark Everingham
Luc Van Gool
Chris Williams
John Winn
Andrew Zisserman
Classification Challenge

- Predict whether at least one object of a given class is present in an image

- Competition 1: Train on the supplied data
  - Which methods perform best given specified training data?

- Competition 2: Train on any (non-test) data
  - How well do state-of-the-art methods perform on these problems?
Evaluation

- **Average Precision [TREC]** averages precision over the entire range of recall
  - Curve interpolated to reduce influence of “outliers”

- A good score requires both high recall **and** high precision

- Application-independent
Participation

- 48 Methods, 20 Groups
- VOC2008: 21 Methods, 11 Groups
- Overwhelmingly “bag of visual words” methods with multiple features e.g. SIFT, color
- Multiple submissions of methods with small variations e.g. different features or classifier architectures
Results: AP by Method and Class

- Only methods in 1st, 2nd or 3rd place by group shown
- Groups: CVC, FIRST/Nikon, NEC/UIUC, UVA/Surrey
Median AP: Best Result by Group
Precision/Recall: Aeroplane (All)

---

**NECUIUC_CDCV (88.1)**
**NECUIUC_CLS-DTCT (88.0)**
**NECUIUC_LL-CDCV (87.7)**
**NECUIUC_LL-CDCV (87.1)**
**CVC_PLUS (86.6)**
**CVC_FLAT-HOG-ESS (86.3)**
**CVC_FLAT (86.3)**
**UVASURREY_TUNECOLORKERNELSEL (85.0)**
**UVASURREY_MKFDAD+ (84.7)**
**UVASURREY_TUNECOLORSPECKD (84.6)**
**UVASURREY_BASELINE (84.1)**
**FIRSTNIKON_AVGSVM (83.8)**
**FIRSTNIKON_BOOSTSVMS (83.5)**
**FIRSTNIKON_AVGSRKDA (83.3)**
**FIRSTNIKON_BOOSTSRKDA (83.0)**
**LIP6_SS-SPK-SVM (80.9)**
**LEOBEN_SCC-200 (80.4)**
**LEOBEN_CHI-SVM-MULT (79.7)**
**LEOBEN_CHI-SVM-MULT-LOC (79.5)**
**LEOBEN_SCC-CLS (79.5)**
**CASIA_SVM-MULTIFEAT (78.2)**
**LIP6_HS-SPK-SVM (77.9)**
**LEOBEN_DENSESIFT (77.0)**
**RTSU_WSF (76.9)**
**MPI_STRUCT (75.9)**
**RTSU_AKF (75.9)**
**RTSU_ASF (75.4)**
**IP_SVM-ROIC (74.6)**
**KERLE_SVM-DENSESIFT (74.2)**
**LIRIS_EER (74.1)**
**LIRIS_BASELINE (73.5)**
**LIG_MRIM-FUSION (71.6)**
**ALCALA_AVW (70.9)**
**ALCALA_LAVW (70.4)**
**LIRIS_SOFT-EER (70.3)**
**LIRIS_SOFT-BASELINE (70.0)**
**UC3M_GEN-DIS (69.9)**
**LIG_MRIM-COLORSIFT (69.5)**
**CNRS_FUSE-KNN-CTS (68.8)**
**LIG_MRIM-VPH (62.0)**
**CNRS_KNN-OHLS (61.5)**
**CNRS_KNN-CT (57.6)**
**CNRS_KNN-GABOR (51.5)**
**HAS_FISHSEIFT-FISHSEG (51.0)**
**TSINGHUA_ALL-SVM-BOOST (45.5)**
**CNRS_KNN-RGB (42.1)**
**TSINGHUA_SVM-SEG-HOG (32.7)**
**FIRST_L2MKL (6.8)**
Precision/Recall: Aeroplane (Top 10 by AP)

Top 10 results by AP

- NECUIUC_CDCV (88.1)
- NECUIUC_CLS-DTCT (88.0)
- NECUIUC_LN-CDCV (87.7)
- NECUIUC_LL-CDCV (87.1)
- CVC_PLUS (86.6)
- CVC.FlatHOG-ESS (86.3)
- CVC.Flat (85.3)
- UVASURREY_TUNECOLORKERNELSEL (85.0)
- UVASURREY_MKFDABOW (84.7)
- UVASURREY_TUNECOLORSPECKDA (84.6)
Precision/Recall: Bicycle (All)

[Diagram showing precision-recall curves for various methods with thresholds and accuracy scores provided for each method.]
Precision/Recall: Bicycle (Top 10 by AP)

Top 10 results by AP:
- NECUIUC_CLS-DTCT (68.6)
- NECUIUC_CDCV (68.0)
- NECUIUC_LN-CDCV (67.8)
- NECUIUC_LL-CDCV (67.4)
- UVASURREY_MKFDABOW (63.9)
- UVASURREY_TUNECOLORKERNELSEL (62.8)
- UVASURREY_TUNECOLOURSPECKDA (62.4)
- CVC_FLAT-HOG-ESS (60.7)
- FIRSTNIKON_AVGSRKDA (59.3)
- FIRSTNIKON BOOSTSRKDA (59.2)
Precision/Recall: Potted plant (Top 10 by AP)

CVC_FLAT-HOG-ESS (40.8)
LEAR_CHI-SVM-MULT-LOC (39.1)
UVASURREY_TUNECOLORKERNELSEL (39.1)
UVASURREY_TUNECOLORSPECFDATA (38.8)
UVASURREY_MKFDABOW (38.5)
CVC_PLUS (37.5)
CVC_FLAT (37.4)
NECUIUC_CLS-DTCT (36.6)
UVASURREY_BASELINE (34.8)
FIRSTNikon_AVGSRKDA (34.2)
- Max AP: 88.1% (aeroplane) ... 40.8% (potted plant)
Statistical Significance

- Friedman/Nemenyi analysis
  - Compare differences in **mean rank** of methods over classes using non-parametric version of ANOVA
  - Mean rank must differ by at least 5.6 to be considered significant (p=0.05)
Ranked Images: Aeroplane

- Class images: Highest ranked
  ![Class images: Highest ranked](image1)

- Class images: Lowest ranked
  ![Class images: Lowest ranked](image2)

- Non-class images: Highest ranked
  ![Non-class images: Highest ranked](image3)

- Context?
Ranked Images: Bicycle

- Class images: Highest ranked

- Class images: Lowest ranked

- Non-class images: Highest ranked

- “Texture”?
Ranked Images: Cat

- Class images: Highest ranked

- Class images: Lowest ranked

- Non-class images: Highest ranked

- “Composition”?
Ranked Images: Chair

- Class images:
  - Highest ranked

- Class images:
  - Lowest ranked

- Non-class images:
  - Highest ranked

- Scene context?
Do these methods have a bias toward larger objects?

- Moderate evidence for some classes e.g. bicycle, car
- Accuracy tends to peak by 5% of image area
AP vs. Object Class Area

- For most classes, correlation with object class area is zero or negative.

- Methods are learning more about context/scene appearance than object appearance?
- Possibility of occlusion negatively effects accuracy?
VOC2008 vs. VOC2009 Test Data

- High correlation, slightly better results on 2009 – “over-fitting”?
- Best methods are better than best 2008 result – better methods and/or advantage of more training data
Prizes

- **Winner:**
  - **NEC/UIUC**
    - Yihong Gong, Fengjun Lv, Jingjun Wang, Chen Wu, Wei Xu, Jianchao Yang, Kai Yu, Xi Zhou, Thomas Huang
    - *NEC Laboratories America; University of Illinois at Urbana-Champaign*

- **Honourable mentions:**
  - **UVA/SURREY**
    - Koen van de Sande, Fei Yan, Atif Tahir, Jasper Uijlings, Mark Barnard, Hongping Cai, Piotr Koniusz, Theo Gevers, Arnold Smeulders, Krystian Mikołajczyk, Josef Kittler
    - *University of Amsterdam; University of Surrey*
  - **CVC**
    - Fahad Shahbaz Khan, Joost van de Weijer, Andrew Bagdanov, Noha Elfiky, David Rojas, Marco Pedersoli, Xavier Boix, Pep Gonfau, Hany salahEldeen, Robert Benavente, Jordi Gonzalez, Maria Vanrell
    - *Computer Vision Centre Barcelona*